Effectiveness of computerised spelling training in children with language impairments: a comparison of modified and unmodified speech input

Dorothy Bishop, Caroline Adams and Annukka Lehtonen
Department of Experimental Psychology, University of Oxford

Stuart Rosen
Department of Phonetics and Linguistics, University College London

This study evaluated a computerised program for training spelling in 8- to 13-year-olds with receptive language impairments. The training program involved children typing words corresponding to pictured items whose names were spoken. If the child made an error or requested help, the program gave phonological and orthographic cues to build up the word’s spelling. Eleven children received this training with ordinary speech, and eleven had the same program but with speech modified to lengthen and amplify dynamic portions of the signal. Nine children were in an untrained control group. Trained children completed between 6 and 29 training sessions each of 15 minutes, at a rate of 3 to 5 sessions per week, with an average of over 1000 trials. Children were assessed before and after training. Trained children learned an average of 1.4 novel spellings per session. The trend was for children presented with modified speech to do less well than those trained with ordinary speech, regardless of whether they had auditory temporal processing impairments. Trained groups did not differ from the untrained control group in terms of gains made on standardised tests of spelling or word and nonword reading. This study confirms the difficulty of training literacy skills in children with severe language impairments. Individual words may be learned, but more general knowledge of rule-based phonological skills is harder to acquire.

Background

Computers have considerable potential for use in training children with learning disabilities. Their advantages are that (i) they make it easy to present children with many repeated exposures to specific educational tasks; (ii) it is possible to give children automated feedback in a game-like format; (iii) the computer can be programmed to respond adaptively to the child’s level of performance, so that training is focused on
materials that are just beyond current competence; (iv) every response made by the child can be logged, giving a detailed record of learning. Nevertheless, one of the most difficult tasks faced by many children with learning disabilities, learning to read, is not so easy to integrate with computer instruction, because computer speech-recognition systems are not yet adequate to the task of reliably judging if a word has been correctly pronounced. Although various ingenious methods of computer-based reading instruction have been developed (Lynch, Fawcett & Nicolson, 2000; Olson & Wise, 1992), spelling is easier to turn into a computerised task. The child can be presented with a spoken word (which a computer can readily deliver) and the program can then judge whether its name has been entered correctly. A spelling task can be used not only to train spelling accuracy, but also to train the component skill of converting from sounds to letters. In principle, it should be possible to train children to learn specific phoneme-grapheme correspondences, knowledge of which should facilitate reading as well as spelling. This is of potential importance, because ability to use phoneme-grapheme conversion rules is often poor in children with learning disabilities, and in some cases may never become fully automatic. This is particularly the case for children with speech and language impairments, many of whom struggle to achieve functional literacy (e.g. Bishop & Adams, 1990).

It has been suggested that one underlying reason for such difficulty may be an auditory processing disorder that has the effect of making it difficult for children to decompose speech into sub-syllabic units. Suppose a child could hear the difference between ‘bug’, ‘beg’, ‘big’ and ‘bag’, but did not appreciate that all these words began and ended with the same sounds; it would not then be obvious that the spellings of these words have letters in common. There are various theoretical accounts of auditory processing disorders that account for children’s literacy problems in this way. The best-known is a theory that attributes language and literacy problems in children to poor auditory temporal resolution, i.e. a difficulty in distinguishing sounds that are brief or occur in rapid succession (see Tallal, 2000, for an overview). Tallal’s auditory processing account has generated particular interest because it provided the motivation for developing an intervention program that was designed to train children’s ability to perceive salient acoustic features of speech. In this program, FastForWord® (FFW), children play computerised games that use speech stimuli that have been modified to lengthen and amplify dynamic portions of the signal. Although FFW was initially designed to remediate spoken language comprehension, the auditory processing theory has been extended to encompass reading difficulties, and the program has subsequently been extended to include games designed to enhance literacy skills. However, there is continuing debate over the efficacy of FFW (see e.g. Cohen et al., forthcoming; Hook, Macaruso & Jones, 2001; Pokorni, Worthington & Jamison, 2004; Troia & Whitney, 2003).

Goals of the current study

The main goal of this study was to consider whether literacy skills could be enhanced in children with language impairments using a computerised training program that taught children correct spellings by breaking up words into graphemes and phonemes. A subsidiary goal was to see whether such training could be made more effective by incorporating the kind of speech modification that is used in FFW. Two forms of the training program were compared: one used ordinary speech; the other used speech modified using the same algorithms as used in FFW. The prediction was that, if speech
modification enhances the child’s ability to segment phonemes, then intervention with modified speech should be superior to intervention with ordinary speech. Both before and after intervention, children’s spelling ability was assessed using a standardised test: the goal was to see whether scores on this primary outcome measure improved. In addition, we assessed ability to read single words and nonwords, predicting that if the training enhanced phonological awareness, then nonword reading in particular might improve. Children’s auditory processing skills were also assessed in this study, so that we could determine whether the intervention with modified speech was particularly effective for those with auditory problems.

Methods

Participants

Children with language impairments aged from 8 to 13 years were recruited from specialist educational placements for children with communication problems, including four special schools and three special classes within mainstream schools. Children with known sensorineural hearing loss or other handicapping conditions were excluded from consideration. Nonverbal ability was assessed, but was not a selection criterion for this study. To be included the child had to score at least 1 SD below the normative mean on at least one of two comprehension measures: the Test for Reception of Grammar-2 (TROG-2: Bishop, 2003), and the Comprehension scale of the Expression, Reception and Recall of Narrative Instrument (ERRNI: Bishop, 2004). All children meeting these criteria for whom parental permission was given were included in the study. Our original goal had been to recruit 24–30 children per intervention group; however, on the basis of unpromising results from an interim analysis, we decided not to implement the intervention at two schools that had planned to administer it during the summer term.

Assessments

Each child was tested before training began (time 1), and again (time 2) after an average of 12 weeks (range 6–21 weeks), after training had been completed for the trained groups. At time 1 children were given an audiometric screening to ensure they could hear pure tones at 500, 1000, 2000 and 4000 Hz in the better ear at a level of 25 dB SPL. Nonverbal ability was measured at time 1 using the Matrix Reasoning test of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). Expressive and receptive oral language skills were measured at time 1 and time 2 using parallel forms of the Expression, Reception, Recall and Narrative Instrument (ERRNI; Bishop, 2004) and Test for Reception of Grammar-2 (TROG-2; Bishop, 2003). Accuracy of speech production was assessed using the Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986): error scores on this test were highly skewed, and so were transformed by taking the logarithm of 1 + error total. Speeded reading of words and nonwords was assessed at times 1 and 2 using forms A and B of the Test of Word Reading Efficiency (TOWRE: Torgesen, Wagner & Rashotte, 1999) Sight Word Efficiency and Phonemic Decoding Efficiency subtests. Spelling was tested on both occasions using the Spelling subtest of the Wechsler Objective Reading Dimensions (Rust, Golombok & Trickey, 1993). For spelling, the same materials were used at time 1 and time 2 because no parallel form was available.
In addition, auditory function was tested at times 1 and 2 using experimental measures described in more detail by Bishop, Adams, Nation and Rosen (forthcoming). There were two non-speech tasks in which the child had to judge which of two glides (i.e. sounds moving up or down in pitch) moved in a different direction from a standard glide. In Version A, task difficulty was adaptively varied by modifying the duration of the glide (maximum 400 ms, minimum 20 ms), and in Version B, difficulty was varied by modifying the frequency excursion of the glide (maximum 500–1500 Hz, minimum 998–1002 Hz). For the purposes of this paper, we shall report data on version A only, as this task measures the kind of auditory temporal processing skills that Tallal’s theory is concerned with. We predicted that children with poor glide-discrimination thresholds should benefit most from modified speech. The speech tasks used minimal pairs of words taken from Adlard and Hazan (1998), which were presented in quiet or noise.

All discrimination tasks adopted a three-interval two-alternative forced-choice paradigm, in which three animals were shown on a computer screen. For each trial, there were three intervals in which an animal moved as a sound was presented. The animal corresponding to the middle interval was located on the screen above the other two animals, and the examiner explained that the task was to work out which of the two other animals (left or right) made the same sound as the central animal. Demonstration was given to ensure the child understood what was required, and testing proceeded only when the child had performed correctly on at least four consecutive trials at the easiest level. The child responded by selecting the left or right-hand animal by pointing to it. The task was self-paced, with the examiner initiating each trial when the child was ready to attend. A correct response resulted in an item being added to a stack of icons on the left of the screen as a cheerful noise was made; an incorrect response led to a ‘sigh’ noise and no additional icon. Items could be repeated at the child’s request.

For the speech discrimination task, there was a fixed number of trials: four practice trials with easy items, followed by 12 test trials with noise and 12 without noise, randomly intermixed. For the glide discrimination tasks the same format was adopted, except that the difficulty level was adaptively varied according to the child’s level of accuracy, using a more virulent PEST algorithm (Findlay, 1978). In this procedure, difficulty is progressively increased until an error is made, at which point the discrimination is made easier; this continues, with difficulty level being adjusted in smaller and smaller steps. Testing continued until an estimate of the threshold level (75% correct) was reached, or 60 trials had been presented.

**Allocation of children to training groups**

Children were assigned code numbers and allocated to one of the three groups, modified speech (M), ordinary speech (S) and untrained control (U), by a person who was unaware of the children’s identity. A minimisation method was adopted: this is a method that ensures that groups do not differ substantially on crucial background variables (Altman et al., 2000). The list of codes for potential participants was rank ordered first by school, then by nonverbal IQ and finally by age, and allocation was then made to the three groups in the order M, S, U, U, S, M, M, S (repeatedly applied) to minimise group differences on these variables and to ensure a roughly 3:3:2 ratio of the three groups. Testers were kept unaware of which training group the child was in.
Spelling training program

A new computer program was developed by the first author to train spelling, using principles of errorless learning (Fillingham et al., 2003). The program was presented as a game, in which the child saw a pictured item and heard its name spoken, and had to fill in a grid by typing letters corresponding to the correct spelling. If the child made an error or pressed a ‘Help’ button, then the program provided help in incremental steps. The first step was to provide the initial spoken sound, e.g. for the word HAT, ‘it begins with /h/’. If this did not lead to success, at the next step, the same information was provided and the correct letter was briefly shown in the grid. If further help was needed, then the program gave the next spoken sound, ‘the sound is æ’; again, if necessary, the program briefly showed the next letter. This process of building up the sounds and letters in the word proceeded until the child typed the correct spelling. Thus, for each word, the child eventually had the experience of correctly typing the name before proceeding to the next trial. For each word completed, the child was rewarded by having an icon added to a stack on the screen, and if the word was spelt correctly without help, then the pictured object spiralled off the screen and a rewarding sound was played. There were 14 levels of the game, as shown in Table 1; these covered a wide range of difficulty so that training could be challenging for children who started the program with a wide range of initial spelling ability. For the higher levels, grapheme-phoneme correspondences were shown

<table>
<thead>
<tr>
<th>Level</th>
<th>Words at this level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: regular CVC with vowel E or A</td>
<td>hen, rat, hat, pan, red, mat, bed, bag, peg, fan</td>
</tr>
<tr>
<td>Level 2: regular CVC with various vowels</td>
<td>dog, bus, pin, zip, cat, gun, fox, sit, jug, bat, sun, cup, tin, web, jam</td>
</tr>
<tr>
<td>Level 3: regular CVC with consonant digraph</td>
<td>hill, ball, sack, fish, king, lock, shell, kick, sock, duck, bell, doll, egg, ring, tack, suck</td>
</tr>
<tr>
<td>Level 4: regular CVC with vowel digraph</td>
<td>sheep, bee, boot, pool, feet, car, bird, boat, book, fork, leaf, owl, saw, soap, nail, bark</td>
</tr>
<tr>
<td>Level 5: regular CVC with long vowel and silent E</td>
<td>bike, cake, gate, kite, tie, shape, chase, nose, pipe, dive</td>
</tr>
<tr>
<td>Level 6: regular CVC with consonant cluster or CH</td>
<td>tree, drum, pram, flag, lamp, bulb, chips, clock, skate, chick, milk</td>
</tr>
<tr>
<td>Level 7: regular CVC with cluster and digraph</td>
<td>bricks, black, scarf, chimp, clown, grape, snail, snake, spoon, swing, tree, track</td>
</tr>
<tr>
<td>Level 8: irregular vowel</td>
<td>bear, door, eye, glove, shoe, vase, watch, pear, grass, key</td>
</tr>
<tr>
<td>Level 9: regular two-syllable</td>
<td>camel, cooker, hippo, jacket, rabbit, rocket, ladder, pecking, finger, devil</td>
</tr>
<tr>
<td>Level 10: irregular or silent consonant</td>
<td>climb, chair, fence, horse, knife, plant, calf, comb, knee, palm, whale</td>
</tr>
<tr>
<td>Level 11: more complex two syllable</td>
<td>apple, baby, below, cushion, guitar, orange, table, turtle, kettle, angel</td>
</tr>
<tr>
<td>Level 12: longer regular word</td>
<td>butterfly, daffodil, dinosaur, hedgehog, kangaroo, octopus, toothbrush, umbrella, transport, lollipop, rainbow</td>
</tr>
<tr>
<td>Level 13: longer words with irregular correspondences</td>
<td>elephant, pineapple, tricycle, telescope, underwear, mountains, cigarette, curtain, pyramid, scissors</td>
</tr>
<tr>
<td>Level 14: longer words including those with longer units such as –tion or –ful</td>
<td>invitation, beautiful, accommodation, language, institution</td>
</tr>
</tbody>
</table>

© United Kingdom Literacy Association 2005
for multi-letter sequences, e.g., the letters SH were entered in the grid, as the computer presented the sentence ‘the sound is /ʃ/’. All the words at a given level were presented, and then any words at that level that had not been spelt correctly without help were repeatedly re-presented until the child was able to type the name without help. Only at this point did the program proceed to present words from the next level. From one session to the next, the program stored information about the child’s level of accuracy so that the next session started at the lowest level at which an error had been made.

All words, word segments and carrier phrases (e.g. ‘it begins with . . .’) were recorded in an anechoic chamber by a young native female speaker of standard Southern British English. In the Modified Speech condition (M), each word, phrase or segment was first run through the algorithm used by FFW, based on descriptions by Nagarajan et al. (1998) and Uchanski, Geers and Protopapas (2002). For the first ten sessions, the maximum level of modification was used, with a smaller degree of modification for the next ten sessions, and no modification for the last five sessions.

Children who were trained on this game were also trained in the same session on a grammar game designed to improve comprehension by having the child move objects on the screen to correspond to spoken commands. This game also lasted 15 minutes, also used ordinary or modified speech to give feedback for groups S and M respectively, and was typically completed in the same session as the spelling game. Results from the grammar game will be presented elsewhere.

Training regime

Training was administered by school staff, starting during March to April 2003. School staff were introduced to the program by the researchers and were given written instructions about how to administer the training, as well as technical support over the telephone. The program was devised so as to require minimal expertise from teachers and to be simple enough for children to quickly learn to run their own sessions. The main role for the teacher was to start the child on each session, and intervene if the child had problems operating the program. Teachers were asked to encourage the child to use the ‘Help’ buttons rather than to provide help directly themselves, so that the computer log of responses would accurately reflect the child’s unaided responses. The goal was for each child to do 20–25 training sessions, each lasting 15 minutes. In general, training sessions took place on consecutive schooldays, but some children had breaks in training for holidays. School staff scheduled the training so as to cause minimal disruption to regular classroom lessons, and speech and language therapy sessions. Children in the untrained control group continued to receive their usual special educational input during the period of the study.

Results

Pre-training profiles

One school withdrew from the trial after experiencing difficulty in running the training programs on their school computers. Five children were absent from school when the testers visited for post-testing and their data will not be considered further. In addition, five children (three from group M and two from group S) completed fewer than six training sessions, and were reclassified into a drop-out group. Ages and test scores of the training groups are shown in Table 2. ANOVA confirmed that the groups did not differ
significantly on any of the pre-training tests, although there was a trend for the drop-out
group to do poorly on most measures.

Auditory processing. Bishop et al. (forthcoming) presented a detailed account of
findings on experimental measures of auditory non-speech and speech discrimination in a
sample of language-impaired children that included those children from the current study
who had non-verbal scaled scores on Matrix Reasoning of 40 or above (i.e. within 1 SD
of the population mean). Compared with a typically developing control group with a
mean age 10.8 years, and mean Matrix Reasoning T-score of 51.8, most children in the
current study performed within normal limits on the auditory processing measures, even
though their Matrix Reasoning was much lower (see Table 2). Nevertheless, three
children in group S, four in group M and three in group U were either unable to score
above chance on the duration discrimination task, or had very high thresholds. Thus these
children had evidence of the kind of auditory perceptual difficulties that motivated
development of the modified-speech intervention.

Amount of training received

Although the goal had been for children to complete 20 sessions, each of 15 minutes,
many children fell short of this goal. The reasons for this included difficulty in scheduling
the training in the context of other classroom activities, reluctance of children to continue
and problems with computer hardware/software. The number of sessions and total
number of training trials are shown for groups M and S in Table 3. The groups did not
differ on either variable, though there was wide variation in both groups: this will be
taken into consideration when examining the outcome of training.

Improvement on the training program

Each word in Table 1 was categorised for each child as (i) not learned (never spelt
without help when first presented in a session); (ii) learned (help needed on first
presentation but spelt without help when first presented in a subsequent session); (iii)
already known (spelt correctly without help on first presentation); (iv) not presented (child did not progress to words of this level during the course of training). Numbers of words attempted, known and learned are shown for each training group in Table 4. These data indicate that the program had some success in training children to spell words whose spellings they had previously not known, with around 20 words being learned on average. As one might expect, the number of words learned was significantly correlated with the number of sessions, Pearson $r = 0.567$, $n = 22$, $p = 0.006$. However, less expected was the finding that this correlation was driven solely by group S, for whom the correlation was 0.88, $n = 11$, $p < 0.001$, whereas in group M the correlation was non-significant, $r = 0.29$, $n = 11$, $p = 0.392$. This difference in size of correlation coefficients is statistically significant at the 0.05 level; $z = 2.15$ (Guilford & Fruchter, 1973). The scatterplot depicting these relationships is shown in Figure 1, together with the trend lines showing the linear regression in each group. Figure 1 suggests that prolonged training is more effective for group S than for group M. A measure of words learned per session was computed by dividing total words learned by number of sessions. Although this did not differ significantly for the two training groups overall (see Table 4), when attention was restricted to those completing 14 or more sessions (6 from group S and 8 from group M), then the difference just reached significance, $F(1, 12) = 4.73$, $p = 0.05$, mean for group $S = 1.56$, $SD = 0.34$, mean for group $M = 0.99$, $SD = 0.57$. This was a post hoc analysis that must be interpreted with caution, but it suggests that insofar as there are differences between training regimes, those in group S may fare better, with the gap from group M widening as training proceeds.

The question arises as to whether the modified speech training is more effective than ordinary speech for children with poor auditory processing. To address this question, the sample was subdivided according to whether the glide discrimination thresholds were within normal limits. Those who had thresholds more than 1.5 SD from a typically-developing mean, or who did not perform above chance on the glide discrimination task, were grouped together as having ‘auditory impairment’. A two-way ANOVA was conducted to assess the effects of training group and auditory impairment on words learned per session. Neither main effect was significant ($F$-ratios $<1$), nor was the interaction, $F(1, 18) = 1.65$, $p = 0.215$. The trend was for children with auditory impairment to do worse with modified than with ordinary speech. Thus the mean words learned per session for the four children with auditory impairment in group M was 0.86 ($SD = 0.54$), whereas for the three children with auditory impairment in group S it was 1.60 ($SD = 0.89$). Means for the children with no auditory impairment were 1.63

### Table 3. Mean number of training sessions, total number of trials completed and highest level attempted for training groups.

<table>
<thead>
<tr>
<th></th>
<th>Group S</th>
<th>Group M</th>
<th>$F(1, 20)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sessions: mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>range</td>
<td>13.91 (7.37)</td>
<td>17.55 (7.27)</td>
<td>1.36</td>
<td>0.258</td>
</tr>
<tr>
<td>trials: mean (SD)</td>
<td>973.45 (606.74)</td>
<td>1216.64 (694.71)</td>
<td>0.76</td>
<td>0.392</td>
</tr>
<tr>
<td>range</td>
<td>210 to 2201</td>
<td>287 to 2742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>highest level: mean (SD)</td>
<td>8.18 (4.00)</td>
<td>8.45 (3.59)</td>
<td>0.03</td>
<td>0.868</td>
</tr>
<tr>
<td>range</td>
<td>3 to 13</td>
<td>3 to 14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(SD = 1.05) for seven children in group M, and 1.43 (SD = 0.55) for eight children in group S.

**Literacy scores at time 2**

Standardised scores on the literacy measures are shown for the three groups in Table 5, together with the change in raw score from time 1 to time 2. Change between time 1 and time 2 was evaluated using the raw change scores, because floor effects on the scaled scores made these relatively insensitive. Two way repeated measures ANOVA was used to assess differences between time 1 and time 2 scores for the three groups. There was a significant effect of test occasion (time 1 versus time 2) on TOWRE nonwords, $F(1, 26) = 5.36, p = 0.029$, but not on TOWRE words. However, this reflected a trend for raw scores to decline with time on the nonword task, contrary to prediction, possibly reflecting greater difficulty of nonwords in form B compared with form A. There was improvement with time on the spelling test, $F(1, 27) = 4.76, p = 0.038$, but note that the
spelling test was the only literacy measure where no parallel form was available, so the same items were used at time 1 and time 2. On no literacy test was there a significant interaction with group (all F ratios < 1). Although there was evidence of modest improvement on retest on the spelling test, this was no different for untrained or trained children. Furthermore, the variance in spelling improvement was as large in untrained as in trained children, which would not be the case if there was a subset of children who had responded positively to the intervention. To see if there was any generalisation from the training to the standardised spelling test, we computed the correlations between number of words learned in training and gain in raw spelling score. This was non-significant (r = 0.073, n = 22, p = 0.747). Thus children who learned more words during the training did not make any more improvement on the spelling test than those who learned few words. The correlation between total number of training sessions and improvement in spelling also fell short of significance, r = 0.37, n = 22, p = 0.091.

Predictors of improvement

We considered whether children’s spelling improvement, either on the standardised test, or within the training task, might be related to language or non-verbal ability, or auditory processing skills. To measure improvement on the training task, we took the mean words learned per session. To simplify the data, the TROG-2 and ERRNI comprehension were averaged to form a receptive language composite, and the other ERRNI indices averaged to give an expressive composite. In no case did the correlation between outcome and predictor variables reach statistical significance (see Table 6). Exploration of correlations within each training group did not materially alter this pattern of results.

Discussion

This study addressed two related questions: first, is it possible to enhance literacy in language-impaired children by using computer-based training of spelling and phonological awareness, and second, does modification of speech input in such a program improve learning in children with auditory processing problems. The brief answer to the first question is that children could be trained to spell new words in the context of a computerised game, but there was no evidence that the training led to more generalised improvements in phonological awareness, insofar as the trained children did no better than untrained children on a spelling test with different words, nor did they
show any improvement in nonword reading. Regarding the second question, we found no evidence of any difference in outcome for children who were trained with ordinary or modified speech. There was a non-significant trend for children with auditory impairments to do worse with modified than with ordinary speech.

In evaluating our data, it is important to bear in mind some limitations of the study. The sample size was small, giving us limited power to detect small changes. Nevertheless, the data did not show any suggestive trends that would encourage us to repeat the study with a larger sample. In arguing that training did not generalise, we relied on standardised literacy tests, and it could be that this was too gross a measure. In future work, it would be useful to devise a test of generalisation that used word sets analogous to those that were trained.

It is possible that more positive findings could have been obtained had we used a longer intervention period, and/or longer training sessions. However, recent studies have obtained results similar to these, with little evidence of any specific improvements in literacy or phonological skills, for children undergoing the more intensive FFW training, where sessions last 90–100 minutes and are typically given five days per week for six weeks (Cohen et al., forthcoming; Hook, Macaruso & Jones, 2001; Pokorni, Worthington & Jamison, 2004; Troia & Whitney, 2003). Furthermore, longer sessions would not be easy to accommodate in the typical school day and there can be difficulties in sustaining children’s motivation over extended sessions. In general children appeared to enjoy the program, and were happy to complete daily 15 minute sessions, but some became frustrated or bored at lack of progress, and our data suggested that those who dropped out tended to be those with poorest language comprehension and literacy skills. It should also be noted that, although every effort was made by school staff to ensure that trained children did not miss crucial activities, time spent on this kind of computerised training

---

Table 6. Spearman correlations (ns) between predictor variable and indices of spelling improvement. 1

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Statistic</th>
<th>Raw gain spelling</th>
<th>Words learned per session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>$r$</td>
<td>$-0.194$</td>
<td>$0.199$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.295$</td>
<td>$0.374$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$31$</td>
<td>$22$</td>
</tr>
<tr>
<td>Matrices ss</td>
<td>$r$</td>
<td>$-0.065$</td>
<td>$-0.093$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.729$</td>
<td>$0.681$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$31$</td>
<td>$22$</td>
</tr>
<tr>
<td>Expressive composite</td>
<td>$r$</td>
<td>$0.150$</td>
<td>$0.042$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.428$</td>
<td>$0.858$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$30$</td>
<td>$21$</td>
</tr>
<tr>
<td>Receptive composite</td>
<td>$r$</td>
<td>$-0.020$</td>
<td>$0.031$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.915$</td>
<td>$0.892$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$31$</td>
<td>$22$</td>
</tr>
<tr>
<td>Articulation (log error score)</td>
<td>$r$</td>
<td>$-0.025$</td>
<td>$0.133$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.894$</td>
<td>$0.556$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$30$</td>
<td>$22$</td>
</tr>
<tr>
<td>Speech discrimination</td>
<td>$r$</td>
<td>$0.103$</td>
<td>$-0.176$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.582$</td>
<td>$0.434$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$31$</td>
<td>$22$</td>
</tr>
<tr>
<td>Duration threshold, glides</td>
<td>$r$</td>
<td>$0.054$</td>
<td>$-0.078$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$0.783$</td>
<td>$0.738$</td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$28$</td>
<td>$21$</td>
</tr>
</tbody>
</table>

---

154 BISHOP, ADAMS, LEHTONEN and ROSEN

© United Kingdom Literacy Association 2005
may reduce the amount of time spent on regular literacy instruction (Troia & Whitney, 2003). As computers become more widespread in society, it is possible that in future this problem could be overcome by having children do such training at home.

**Future directions**

This study raises a host of questions for future research about how rates of learning might be improved.

1. **Length of training sessions.** It is usually assumed that ‘more is better’, and that if training is to be effective, it needs to involve regular lengthy sessions. However, such an approach may be counterproductive, if it leads to children getting overwhelmed or bored. Our results suggest that it might be better to adopt a ‘little and often’ approach.
2. **Variety in the training set.** The early levels of our training program used word sets with related properties, but it would be possible to make these even more uniform, e.g. by using only rhyming regular word sets, such as CAT, MAT, BAT. A potential drawback of reducing variety is that children may become bored, or that they may learn to apply a specific spelling pattern, such as final AT within a training session, without appreciating its phonological significance.
3. **Reinforcing correct spellings of misspelt words.** Our program worked by presenting all the words at a given level and then returning to re-present any word that was not spelled correctly when first presented. An alternative method would be to immediately re-present any word that was not spelled correctly, so reinforcing the correct spelling right away. Again, one has to weigh up the potential benefits of such focused training against the likelihood that it could increase the child’s frustration and fatigue if the same word keeps recurring. Note too that in the course of training children had some trials that involved spelling words that they already knew: this might help maintain their motivation, but it could be argued that it would be preferable to focus only on words that were not known.
4. **Single word training.** One way of dealing with points 1 to 3 would be to train the spelling of just one new word in any one training session: there would be no set time limit – the word would simply be repeated until the child spelt it correctly without help. Although this may seem a rather limited approach to training, it is worth considering, given that children in the current study learned, on average, just one word every ten minutes with more varied training. It seems possible that learning of a single word might be achieved more quickly if an entire session was focused to that end.
5. **Introducing irregular words.** The structure of English orthography raises special difficulties because one has to decide whether to teach only regular grapheme-phoneme correspondences, to simplify the learning task, or whether this could be counterproductive, given that many words do not obey the rules. In the current programme, the early levels included only regular words, but irregular words were introduced subsequently. It is possible that children would have learned phonological skills that generalised to other words if training had focused solely on regular words. Nevertheless, irregular words do contain regular grapheme-phoneme correspondences (e.g. G, L and V in ‘glove’), and these were reinforced in the training program.
6. **Integration of computerised training with other classroom activities.** The computerised training used in this study was not linked to other educational activities in the children’s classrooms. This allowed us to evaluate the impact of the computerised
training without any confounding influences. However, Nicolson, Fawcett and Nicolson (2000) argued that ‘to be viable in schools, the computer must be an integral part of the teaching process rather than an unintegrated afterthought’ (p. 196). It could be argued that better gains in spelling ability might be achieved if computerised training sessions were used to reinforce classroom-based learning, so that the child had opportunities to read and write the same targeted words in more naturalistic contexts.

Concluding comment

Torgesen (2000) noted that studies of intervention for reading disabilities in the mainstream school setting reliably find a small core of children who fail to respond to phonologically based interventions. He termed these ‘treatment resisters’, and noted that they represent about 2–6% of all children, which is a similar proportion to that referred for special educational services. Nicolson et al. (2002) found that children who were identified as at risk of dyslexia were the least likely to make gains with a reading intervention program, regardless of whether a computerised or traditional method of training was used. It seems likely that low level of oral language skill may be a risk factor for poor response to intervention. In line with this, Whiteley, Smith and Connors (under review) found that weak vocabulary skills characterised children who failed to benefit from intensive phonological interventions. Furthermore, in one of the few well-controlled studies to focus on phonological awareness training in children with receptive-expressive language impairments, Pokorni, Worthington and Jamison (2004) found no improvement in reading ability after intensive training (three hours per day for 20 days) on either Fast Forward, or on one of two other commercially available training programs. Our sample was recruited from children who had severe and persistent language impairments that affected both receptive and expressive language skills. Many of them also performed below age level on a test of nonverbal ability. Most of these children had been in full-time specialist educational placements for some years, and had poor levels of literacy despite intensive and expert intervention. The current study confirms that word spellings can be trained with such children, but it is likely to be a slow process.

Acknowledgements

We thank the staff and pupils of the following schools for their generous assistance with this study: Dawn House School, Alderwasley Hall, Gap House School, The Link Primary School, Meath School, Aston Clinton Combined School, Moor House School, Rose Hill First School and Penhill Primary School. Dawn House School and Meath School are ICAN schools. ICAN is the national educational charity for children with speech and language impairments. This research was supported by a programme grant from the Wellcome Trust.

References


**Address for correspondence:** Dorothy Bishop, Department of Experimental Psychology, University of Oxford, South Parks Road, OX1 3UD, UK. E-mail: dorothy.bishop@psy.ox.ac.uk

© United Kingdom Literacy Association 2005